

Title	A Study of Software Project Checkup Based on Data Mining Tequniques
Author(s)	出張, 純也
Citation	
Issue Date	
oaire:version	
URL	<a href="https://hdl.handle.net/11094/59314">https://hdl.handle.net/11094/59314</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">https://www.library.osaka-u.ac.jp/thesis/#closed</a> 大阪大学の博士論文について <a href="#">ご参照</a> ください。

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏 名	出 張 純 也
博士の専攻分野の名称	博 士 (情報科学)
学 位 記 番 号	第 2 5 2 8 4 号
学 位 授 与 年 月 日	平成 24 年 3 月 22 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 情報科学研究科情報システム工学専攻
学 位 論 文 名	A Study of Software Project Checkup Based on Data Mining Tequiques (データマイニング技術を応用したソフトウェアプロジェクト診断につ いての研究)
論 文 審 査 委 員	(主査) 教 授 菊 野 享 (副査) 教 授 楠 本 真二 教 授 尾 上 孝雄 准教授 土 屋 達弘

## 論 文 内 容 の 要 旨

ソフトウェア開発現場では、開発期間の短縮、信頼性向上、コストの削減など、様々な要求が増加している。こうした状況下で、プロジェクトは混乱状態と呼ばれる状態に陥ることがある。混乱状態とは、プロジェクトの進捗状況が誰にも把握できないような状態のことである。混乱状態となったプロジェクトでは、コスト超過、品質低下、納期遅延など様々な問題が発生する。そのため、混乱状態の回避は急務である。

本論文では、混乱状態を回避するためのプロジェクト管理の方法として「プロジェクト診断」を提案している。この診断は、プロジェクトを「簡易診断モード」と「詳細診断モード」の2状態に分けて管理するものである。

簡易診断モードでは、プロジェクトは少数のメトリクスを収集し、その値に基づいてプロジェクトの予測を行う。プロジェクトが失敗すると予測された場合には、プロジェクトは詳細診断モードに移行する。本論文では、簡易診断モードで利用するメトリクスの選択手法とプロジェクト予測手法について提案している。提案法を利用した結果、7つのメトリクスを用いて82.8%の精度が達成できた。

詳細診断モードでは、プロジェクトは可能な限りメトリクスを収集し、その値に基づいてプロジェクトの予測を行う。ここでプロジェクトが成功すると予測されると、プロジェクトはそのまま開発を実行するが、プロジェクトが失敗すると予測された場合にはプロジェクトの改善を行う必要がある。本論文では、詳細診断モードで利用する予測手法についても提案している。提案法を利用した結果、69個のメトリクスを用いて90%以上の精度が達成できている。

さらに、本論文では詳細診断モードでプロジェクトの改善案を提案するための手法として、プロジェクトの特定のリスクに対して改善案を提案できる「改善ルール」を提案している。ケーススタディでは、11の改善ルールを導出する事ができた。

本論文は次のような構成になっている。まず、1章では本論文の背景と概要を述べる。2章では、本論文で利用しているソフトウェア開発プロジェクトデータと、データマイニ

ング技術について説明する。3章では、提案するプロジェクト診断の概要を述べる。4章では、簡易診断モードでのメトリクス選択手法と予測手法について説明する。5章では詳細診断モードでの予測手法について述べる。6章では、リスク要因とその改善案の抽出手法について述べる。7章では、本論文のまとめと今後の課題について述べる。

## 論 文 審 査 の 結 果 の 要 旨

近年のソフトウェア開発においては、ソフトウェア品質、開発コスト、納期を計画通りに達成することを目指して、定量的なプロジェクト管理が行われている。ところが、プロジェクトはしばしば混乱状態と呼ばれる状態に陥り、最終的に失敗プロジェクトとなってしまう。一度、混乱状態になると、プロジェクトの進捗状況が誰にも把握できなくなり、コスト超過、品質低下、納期遅延など様々な問題が発生するので、混乱状態の回避を実現する新たなプロジェクト管理法の開発が急務となっている。

本論文では、そうしたプロジェクト管理の新たな取組みとして「2段階のプロジェクト診断」を考案し、そのための要素技術の開発を行っている。考案したプロジェクト診断は、いわゆる人間ドックによる精密検査と異常ありの場合の再検査をヒントにしており、「簡易診断モード」と「詳細診断モード」の2段階に分けて診断する。研究の実用性を考慮して、本論文ではIPA/SECが収集・発表しているデータ白書の公表データを利用している。そのため、メトリクスの数は多いが、収集データに欠損が多く、しかも成功と失敗プロジェクトのデータ数の割合に偏りが大きい。こうした利用データの特性の克服も大きな課題である。

先ず、簡易診断モードでは、少数のメトリクスデータに基づいてプロジェクトの成功と失敗を予測する方針をとっている。具体的には、欠損率とその閾値の採用によって欠損の多い大量のメトリクスを削除した後に、相関ルールマイニングを適用して、最終的には7つのメトリクスに絞り込んでいる。更に7つのメトリクスを利用したバイズ識別器に基づく予測モデルを作成し、82.8%の精度を達成した。

一方、詳細診断モードでは、なるべく多くのメトリクスデータを利用する方針の下で、プロジェクトの成功と失敗を予測する。なお、失敗すると予測された場合には、プロジェクトの改善を行う必要がある。具体的には、モデルで利用する成功と失敗プロジェクトのデータ数のバランスを配慮しつつ、欠損の多いプロジェクトを段階的に削除して69個のメトリクスに絞り込んでいる。引き続いて、バイズ識別器に基づく予測モデルを作成し、90%以上の精度を達成した。更に、成功プロジェクトと失敗プロジェクトに共通に発生している事象に注目し、これらの類似点や差分の計算から、最終的に11個の改善ルールを導出した。

以上のように、本論文は、ソフトウェア開発プロジェクト管理の新たな取組みを確立しており、プロジェクトの効率的な管理に貢献するところが大きい。よって、博士(情報科学)の学位論文として価値のあるものと認める。

氏 名	畑 秀 明
博士の専攻分野の名称	博 士 (情報科学)
学 位 記 番 号	第 2 5 2 8 5 号
学 位 授 与 年 月 日	平成 24 年 3 月 22 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 情報科学研究科情報システム工学専攻
学 位 論 文 名	Fault-prone Module Prediction Using Version Histories (履歴情報を用いたフォールトブローンモジュール予測に関する研究)
論 文 審 査 委 員	(主査) 教 授 菊 野 亨 (副査) 教 授 楠 本 真 二 教 授 尾 上 孝 雄 准教授 土 屋 達 弘

## 論文内容の要旨

ソフトウェア開発やメインテナンスにおいて、テストやレビューといった品質確保の活動は、実用的な時間とコストで効果的に行うことが期待される。このため、重点的に品質確保の工数をかけるべき、フォールトブローン（不具合の潜在が疑われる）モジュールを早期に予測することが求められる。近年、ソフトウェアリポジトリからのデータマイニングによって得られる開発履歴メトリクスを用いたフォールトブローンモジュール予測が注目されている。本研究では、2つの手法を提案して現状の課題に取り組んだ。

多くの研究が行われているフォールトブローンモジュール予測であるが、適用が容易でないことが実用化の障壁として挙げられる。特に予測モデル構築に必要な開発履歴メトリクスの計測は、実際の開発に用いられている個別のソフトウェアリポジトリに対応したマイニングツールの開発が必要であり、容易ではない。我々はテキストマイニングに基づく予測手法を提案した。これはソースコードをテキストとして扱い、トークンの出現回数を計測するだけでよいため適用が容易である。オープンソースソフトウェアへの適用実験から、従来の複雑度メトリクスや開発履歴メトリクスに基づく予測モデルと比べても、テキストマイニングに基づく予測モデルは高い精度で予測ができることを確認した。

これまでの開発履歴メトリクスに基づいたフォールトブローンモジュール予測は、モジュールが主にファイルレベルであった。予測されたモジュールを調査する工数を考慮した場合、モジュールのサイズはより小さいことが望ましい。しかし、細粒度なモジュールである、メソッドに対するFault-proneモジュール予測では開発履歴の取得が難しい。この問題に対して、我々は細粒度版管理システム Historage を開発し、メソッドレベルの開発履歴取得を可能にした。このHistorageを用いて計測した開発履歴メトリクスから、メソッドレベルのフォールトブローンモジュール予測を行い、工数を考慮した評価を行った。オープンソースソフトウェアへの適用実験から、同程度の工数で、ファイルレベルと比べて多くのフォールトを調査できることを示した。

開発されるソフトウェアプロダクトの品質確保は極めて重要であり、テストやレビューといった活動によって実現が図られる。しかし現状では、許される開発時間、コスト等への制約が厳しいために十分な活動が展開できず、品質の確保は必ずしも達成されているとは言えない状況にある。このため、フォールトブローン（つまり、フォールトの潜在が疑われる）モジュールを予測して、そのモジュール群に重点的に工数をかけることで作業の効率化を図る開発方法に関する研究が注目されている。

具体的には、ソフトウェアリポジトリに蓄積されている開発データにデータマイニング手法を適用して開発履歴メトリクスを求め、それを用いてフォールトブローンモジュールを予測する多くの試みが報告されている。本論文では、新たに2つの予測手法を提案し、オープンソースソフトウェアへの適用実験を通じて、その有効性を確認している。

まず、フォールトブローンモジュール予測手法そのものについては既に多くの研究があるが、どれも適用が容易でないために未だ実用化には至っていない。特に、予測モデル構築に必要な開発履歴メトリクスの獲得に困難がある。そこで、本論文ではテキストマイニングに基づく予測手法を提案した。ソースコードをテキストとして扱い、トークンの出現回数から予測モデルを構築する。そのため予測モデルの構築を高速に、かつ容易に実行できる。またオープンソースソフトウェアに対する適用実験から、従来の予測モデルと比べても、本手法は高い予測精度が達成できることを確認した。

次に、開発履歴メトリクスに基づいたフォールトブローンモジュール予測に関する研究はこれまでは主として、ファイルレベルでの予測であった。そのため、実際にテストやレビューを行うには、フォールトを含むメソッドをファイルの中から人手で探索する必要があった。予測結果を実際開発で利用しようとするなら、予測するモジュールのサイズは小さいことが望ましい。本論文では、この問題に対処するために細粒度版管理システムHistorageを開発し、メソッドレベルでの開発履歴メトリクス取得とフォールトブローンモジュール予測を可能にした。オープンソースソフトウェアへの適用実験から、従来のファイルレベルでの方法と比べ、より多くのフォールトを検出できることを示した。しかも、そのためのオーバーヘッドは従来法とほぼ同程度であった。

以上のように、本論文はフォールトブローンモジュール予測の新たな手法を提案し、ソフトウェア工学研究に対して大きな貢献を与えるものであり、博士(情報科学)の学位論文として価値のあるものと認める。